

19 BUNDESREPUBLIK
DEUTSCHLAND



DEUTSCHES
PATENTAMT

12 Übersetzung der
europäischen Patentschrift

27 EP 0 372 734 B1

10 DE 689 13 669 T 2

61 Int. Cl. 5:
G 10 L 5/04

21 Deutsches Aktenzeichen:	689 13 669.2
66 Europäisches Aktenzeichen:	89 311 830.7
68 Europäischer Anmeldetag:	15. 11. 89
67 Erstveröffentlichung durch das EPA:	13. 6. 90
67 Veröffentlichungstag der Patenterteilung beim EPA:	9. 3. 94
47 Veröffentlichungstag im Patentblatt:	21. 7. 94

30 Unionspriorität: 32 33 31
23.11.88 US 275581

73 Patentinhaber:
Digital Equipment Corp., Maynard, Mass., US

74 Vertreter:
Betten, J., Dipl.-Ing.; Resch, M., Dipl.-Phys.,
Pat.-Anwälte, 80489 München

84 Benannte Vertragsstaaten:
AT, BE, CH, DE, ES, FR, GB, GR, IT, LI, LU, NL, SE

72 Erfinder:
Vitale, Anthony John, Northborough Massachusetts
01532, US; Levergood, Thomas Mark, Bellingham
Massachusetts 02019, US; Conroy, David Gerard,
Maynard Massachusetts 01754, US

64 Namensausprache durch einen Synthetisator.

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II 53 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patentamt inhaltlich nicht geprüft.

DE 689 13 669 T 2

DE 689 13 669 T 2

Namenaussprache durch einen Synthesizer

5 Die vorliegende Erfindung betrifft eine Text-zu-Sprache-Umsetzung durch einen Computer, und insbesondere das korrekte Aussprechen von echten Namen von Text.

10 Die Namenaussprache kann auf dem Gebiet des Außendienstes innerhalb der Telefonindustrie und der Computerindustrie verwendet werden. Sie wird auch in größeren Firmen gefunden, die eine umgekehrte Verzeichnis-Hilfe (Nummer zu Name) haben, und auch in Text-Nachrichtensystemen verwendet, wo das letzte Namensfeld eine gemeinsame Einheit ist.

15 Es gibt viele Geräte, die kommerziell erhältlich sind und die amerikanisches Englisch als Sprache durch Computer synthetisieren. Eine der für die Sprach-Synthese gesuchten Funktionen, die spezielle Probleme aufweist, ist die Aussprache einer unbegrenzten Anzahl von ethnisch unterschiedlichen Nachnamen. Aufgrund der extrem großen Anzahl von unterschiedlichen Nachnamen in einem ethnisch diversifizierten Land wie die Vereinigten Staaten kann die Aussprache eines Nachnamens unter Verwendung anderer Sprachausgangstechnologien wie zum
20 Beispiel Audioband oder digital gespeicherter Sprache
25 momentan praktisch nicht realisiert werden.

Es gibt typischerweise eine umgekehrte Relation zwischen der Aussprachegenauigkeit eines Sprachsynthesizers in seiner
30 Quellensprache und der Aussprachegenauigkeit des gleichen Synthesizers in einer zweiten Sprache. Die Vereinigten Staaten sind ein ethnisch heterogenes und diversifiziertes Land mit Namen, die von Sprachen abgeleitet sind, die von den gemeinsamen indo-europäischen Sprachen wie zum Beispiel Fran-
35 zösisch, Italienisch, Polnisch, Spanisch, Deutsch, Irisch usw. bis hin zu exotischeren Sprachen wie zum Beispiel Japanisch, Armenisch, Chinesisch, Arabisch und Vietnamesisch reichen. Die Aussprache von Nachnamen von den verschiedenen ethnischen Gruppen stimmt nicht mit den Regeln des

standartisierten amerikanischen Englisch überein. Zum Beispiel werden die meisten deutschen Namen auf der ersten Silbe betont, wohingegen japanische und spanische Namen dazu neigen, auf der vorletzten betont zu werden, und französische Namen eine endständige Betonung haben. Ähnlich wird die orthographische Sequenz CH als [tʃ] in englischen Namen (z.B. CHILDERS), als [ʃ] in französischen Namen wie z.B. CHARPENTIER, und als [k] in italienischen Namen wie z.B. BRONCHETTI ausgesprochen. Menschliche Sprecher erzeugen oft die richtige Aussprache durch "Wissen" der Ursprungs-Sprache des Namens. Das Problem, das durch einen Sprachsynthesizer gegeben ist, besteht im Sprechen dieser Namen unter Verwendung der richtigen Aussprache, denn da Computer nicht den ethnischen Ursprung des Namens "wissen", ist die Aussprache oft falsch.

Ein System ist im Stand der Technik vorgeschlagen worden, in dem ein Name zuerst mit einer Anzahl von Einträgen in einem Verzeichnis verglichen wird, das die bekanntesten Namen aus einer Anzahl von unterschiedlichen Sprachgruppen enthält. Jeder Verzeichnis-Eintrag enthält eine orthographische Form und ein phonetisches Äquivalent. Wenn eine Übereinstimmung auftritt, wird das phonetische Äquivalent zu einem Synthesizer gesendet, der es in eine hörbare Aussprache für diesen Namen umsetzt.

Wenn der Name nicht im Verzeichnis gefunden wird, verwendet das vorgeschlagene System ein statistisches Trigramm-Modell. Diese Trigramm-Analyse umfaßt das Ermitteln einer Wahrscheinlichkeit dafür, daß jeweils eine Drei-Buchstaben-Folge (oder ein Trigramm) in einem Namen mit einer Etymologie verbunden ist. Wenn das Programm ein neues Wort empfangen hat, wurde eine statistische Formel angewendet, um für jede Etymologie eine Wahrscheinlichkeit zu ermitteln, die auf jeder der Drei-Buchstaben-Folgen (Trigramme) in dem Wort basiert.

Das Problem bei dieser Lösung ist die Genauigkeit der Trigramm-Analyse. Das liegt daran, daß die Trigramm-Analyse nur

eine Wahrscheinlichkeit berechnet, und wenn alle Sprachgruppen als möglicher Kandidat für die Ursprungs-Sprachgruppe eines Wortes betrachtet werden, ist die Genauigkeit der Auswahl der Ursprungs-Sprachgruppe des Wortes nicht so hoch, als wenn
 5 weniger mögliche Kandidaten vorhanden sind.

Gemäß einem Aspekt der vorliegenden Erfindung wird ein Verfahren zum positiven Identifizieren oder Eliminieren einer Sprachgruppe als eine Ursprungs-Sprachgruppe für ein gegebenes Wort geschaffen, wobei das Verfahren aufweist:
 10 Vergleichen von Unterketten aus Graphemen eines Eingangswortes mit einem gespeicherten Satz von Filterregeln, bis entweder eine Übereinstimmung einer der Unterketten mit einer der Filterregeln positiv eine Sprachgruppe
 15 identifiziert oder bis irgendeine Sprachgruppe eliminiert wird, wenn eine Übereinstimmung einer der Unterketten mit einer der Filterregeln anzeigt, daß eine Sprachgruppe von der Betrachtung als eine Ursprungs-Sprachgruppe für das Eingangswort eliminiert wird; und
 20 Erzeugen einer Liste möglicher, nicht eliminierter Ursprungs-Sprachgruppen, wenn keine Sprachgruppe positiv als die Ursprungs-Sprachgruppe identifiziert wird, oder anzeigen der Ursprungs-Sprachgruppe, wenn die Ursprungs-Sprachgruppe positiv identifiziert wird.

25 Gemäß einem anderen Aspekt der vorliegenden Erfindung wird ein Verfahren zum Erzeugen korrekter Phoneme für ein eingegebenes Eingangswort gemäß einer Sprachgruppe von Ursprüngen des Eingangsworts geschaffen, wobei das Verfahren aufweist:
 30 Filtern des Eingangsworts in einem Filter, um eine Ursprungs-Sprachgruppe für das Eingangswort zu identifizieren oder um zumindest eine Ursprungs-Sprachgruppe für das Eingangswort zu eliminieren;
 35 Senden des Eingangsworts und einer Sprach-Kennzeichnung, die eine Ursprungs-Sprachgruppe für das Eingangswort anzeigt, von dem Filter zu einem Buchstabe-zu-Klang-Modul, der Buchstabe zu Klangregeln enthält, wenn der Filter positiv eine Ursprungs-Sprachgruppe für das Eingangswort identifiziert;

Senden des Eingangsworts und jeder nicht eliminierten Sprachgruppe von dem Filter aus zu einem Graphem-Analysierer, wenn eine Ursprungs-Sprachgruppe für das Eingangswort nicht positiv durch den Filter identifiziert wird;

5 Erzeugen einer wahrscheinlichsten Ursprungs-Sprachgruppe für das Eingangswort, indem Grapheme in dem Eingangswort analysiert werden;

Senden des Eingangswortes und der wahrscheinlichsten Ursprungs-Sprachgruppe zu einem Untersatz des Buchstabe-zu-Klang-Moduls entsprechend der wahrscheinlichsten Sprachgruppe;

Erzeugen von segmentartigen Phonemen für das Eingangswort in dem Untersatz des Buchstabe-zu-Klang-Moduls;

15 Senden der segmentartigen Phoneme und Sprach-Kennzeichnung von dem Buchstabe-zu-Klang-Modul zu einem Betonungs-Zuordnungsabschnitt;

Erzeugen einer Betonungs-Zuordnungs-Information für das Eingangswort in dem Betonungs-Zuordnungsabschnitt; und

20 Senden der segmentartigen Phoneme und der Betonungs-Zuordnungs-Information zu einer Sprachrealisierungseinheit.

Gemäß diesem Aspekt wird auch eine Vorrichtung zum positiven Identifizieren oder Eliminieren einer Sprachgruppe als eine Ursprungs-Sprachgruppe für ein gegebenes Wort geschaffen, die aufweist:

25 einen Filterregel-Speicher, der einen Satz mit Filterregeln speichert, wobei ein erster Untersatz der Filterregeln positiv eine Sprachgruppe identifiziert und wobei ein zweiter Untersatz der Filterregeln eine Sprachgruppe eliminiert;

30 einen Vergleicher, der Unterketten der Grapheme eines Eingangsworts mit dem ersten Untersatz und dem zweiten Untersatz der Filterregeln vergleicht, bis eine Übereinstimmung einer der Unterketten mit einer des ersten Untersatzes von Filterregeln positiv eine Sprachgruppe identifiziert oder eine

35 Sprachgruppe eliminiert, wenn eine Übereinstimmung einer der Unterketten mit einer des zweiten Untersatzes der Filterregeln anzeigt, daß eine Sprachgruppe von der Betrachtung als eine Ursprungs-Sprachgruppe für das Eingangswort eliminiert ist; und

einen Ausgang, der eine Liste möglicher Ursprungs-Sprachgruppen erzeugt, wenn keine Sprachgruppe als Ursprungs-Sprachgruppe positiv identifiziert wird, und der eine Anzeige der Ursprungs-Sprachgruppe erzeugt, wenn die Ursprungs-Sprachgruppe positiv identifiziert wird.

Die vorliegende Erfindung löst das obenstehende Problem, indem sie die Genauigkeit der Trigramm-Analyse verbessert. Dies wird bewerkstelligt, indem ein Filter bereitgestellt wird, der entweder positiv eine Sprachgruppe als eine Ursprungs-Sprachgruppe identifiziert, oder der eine Sprachgruppe als eine Ursprungs-Sprachgruppe für ein gegebenes Eingangswort eliminiert. Das Filterungsverfahren gemäß der vorliegenden Erfindung umfaßt das Identifizieren oder Eliminieren einer Sprachgruppe als eine Ursprungs-Sprachgruppe für ein Eingangswort gemäß einem gespeicherten Satz von Filterregeln. Der Schritt des Identifizierens oder Eliminierens einer Sprachgruppe enthält das Durchführen einer erschöpfenden Durchsuche des Regelsatzes, indem eine Rechts nach Links-Abtastung verwendet wird. Die Sprachgruppen werden eliminiert, wenn eine Übereinstimmung einer dieser Unterketten mit einer der Filterregeln anzeigt, daß eine Sprachgruppe von der Betrachtung als Ursprungs-Sprachgruppe für das Eingangswort eliminiert werden sollte. Dies wird ausgeführt, bis eine Übereinstimmung einer der Unterketten mit einer der Regeln positiv eine Sprachgruppe identifiziert. Wenn keine Sprachgruppe positiv als eine Ursprungs-Sprachgruppe identifiziert wird, nachdem alle Unterketten für ein gegebenes Eingangswort verglichen wurden, wird eine Liste möglicher Ursprungs-Sprachgruppen aufgestellt. Dieses Filterverfahren erzeugt auch eine positiv identifizierte Sprachgruppe des Originals, wenn eine positive Identifikation gegeben ist.

Die Vorteile des Verwendens eines Filters vor der Trigramm-Analyse enthalten das Vermeiden unnötiger Trigramm-Analyse, wenn Filterregeln positiv eine Sprachgruppe als Ursprungs-Sprachgruppe identifizieren können. Wenn keine Sprachgruppe positiv identifiziert werden kann, reduziert das

Filterungsverfahren die Änderungen einer inkorrekten Annahme, die bei der Trigramm-Analyse gemacht wird, indem die Anzahl der möglichen in Betracht kommenden Sprachgruppen als Ursprungs-Sprachgruppe reduziert wird. Durch die Eliminierung
 5 einiger Sprachgruppen ist die Identifikation einer Ursprungs-Sprachgruppe genauer, wie obenstehend diskutiert wurde.

Die Erfindung enthält auch ein Verfahren zum Erzeugen korrekter Phoneme für ein gegebenes Eingangswort gemäß der Ursprungs-Sprachgruppe des Eingangsworts. Dieses Verfahren umfaßt das Durchsuchen eines Verzeichnisses nach einem Eintrag, der einem Eingangswort entspricht, wobei jeder Eintrag ein Wort und Phoneme für dieses Wort enthält. Dieser Eintrag wird dann zu einer Sprachrealisierungseinheit für Aussprache
 10 gesendet, wenn die Verzeichnis-Durchsuchung einen Eintrag aufdeckt, der dem Eingangswort entspricht. Das Eingangswort wird zu einem Filter gesendet, wenn das Eingangswort nicht einen entsprechenden Eintrag in dem Verzeichnis hat.

Der nächste Schritt beim Verfahren umfaßt das Filtern, um eine Ursprungs-Sprachgruppe für das Eingangswort zu identifizieren oder um zumindest eine Ursprungs-Sprachgruppe für das Eingangswort zu eliminieren. Wenn der Filter positiv eine Ursprungs-Sprachgruppe für das Eingangswort identifiziert, werden das Eingangswort und ein Sprach-Kennzeichen, das eine Ursprungs-Sprachgruppe für das Eingangswort angibt, von dem Filter aus zu einem Buchstabe-zu-Klang-Modul gesendet. Wenn eine Ursprungs-Sprachgruppe nicht positiv durch den Filter identifiziert wird, werden das Eingangswort und alle Sprachgruppen, die nicht eliminiert wurden, von dem Filter aus zu
 20 einem Trigramm-Analysierer gesendet.

Die am meisten wahrscheinliche Ursprungs-Sprachgruppe für das Eingangswort wird erzeugt, indem Trigramme analysiert werden, die in dem Eingangswort auftreten. Diese am meisten wahrscheinliche Ursprungs-Sprachgruppe, die durch die Trigramm-Analyse erzeugt wird, wird zusammen mit dem Eingangswort zu einem Untersatz von Buchstaben-zu-Klang-Regeln gesendet, die mit der wahrscheinlichsten Sprachgruppe korrespondieren. Pho-
 35

neme werden für das Eingangswort gemäß dem entsprechenden Untersatz der Buchstabe-zu-Klang-Regeln erzeugt.

Die Erfindung erstreckt sich in allen Punkten auch auf ein
 5 Verfahren und eine Vorrichtung zur Sprach-Synthese, die die
 obenstehenden Merkmale umfaßt. Die Sprach-Synthese kann eine
 Sprach-Realisierung enthalten, die dafür ausgelegt ist, das
 Wort gemäß der bestimmten Sprache auszusprechen.

10 Das IBM TECHNICAL DISCLOSURE BULLETIN, Vol. 27, Nr. 7A, Dezember 1984, Seite 3681, New York, US; P.S. COHEN usw.:
 "Method for improving spelling-to-sound rules for speech
 synthesis using algorithmically deduced etymologies" offenbart ein Verfahren, das eine positive Identifikation verwendet.
 15

Die vorliegende Erfindung kann auf verschiedene Arten in die Praxis umgesetzt werden, wobei eine von ihnen nachfolgend beispielhaft mit Bezug auf die beiliegenden Zeichnungen beschrieben wird, in denen:
 20

Figur 1 ein logisches Blockdiagramm der Sprachidentifikations- und der Phonem-Realisierungs-Module illustriert; und
 25

Figur 2 ein logisches Blockdiagramm eines Namen-Analyse-Systems zeigt, das den Sprachgruppenidentifikations- und Phonem-Realisierungs-Modul der Figur 1 enthält und in Übereinstimmung mit der vorliegenden Erfindung aufgebaut ist.
 30

Figur 1 ist ein Diagramm, das verschiedene Logikblöcke der vorliegenden Erfindung illustriert. Die physikalische Ausführungsform des Systems kann durch einen kommerziell erhältlichen Prozessor realisiert werden, der logisch wie gezeigt ausgelegt ist.
 35

Ein Name, der ausgesprochen werden soll, wird als Eingang angenommen. Das Suchen wird innerhalb von Einträgen in einem Verzeichnis 10 für diesen Eingangsnamen ausgeführt. Jeder

Verzeichnis-Eintrag hat einen Namen und Phoneme für diesen Namen. Ein semantisches Zeichen identifiziert das Wort als Namen.

- 5 Eine Durchsuchung für einen Eingangsnamen, der einem Eintrag in dem Verzeichnis 10 entspricht, resultiert in einem Treffer. Das Verzeichnis 10 sendet dann sofort den Eintrag (Name und Phoneme) zu einer Sprachrealisierungseinheit 50, die den Namen gemäß den Phonemen ausspricht, die in dem Eintrag enthalten sind. Der Aussprachevorgang für dieses Eingangswort
10 würde dann abgeschlossen sein.

- Ein Verzeichnis-Fehlgriff (dictionary miss) tritt auf, wenn es keinen entsprechenden Eintrag für den Eingangsnamen in dem
15 Verzeichnis 10 gibt. Um die richtige Aussprache bereitzustellen, versucht das System, die Ursprungs-Sprachgruppe des Eingangsnamens zu identifizieren. Dies wird ausgeführt, indem einem Filter 12 der Eingangsname zugesendet wird, der in dem Verzeichnis 10 fehlt. Der Eingangsname wird
20 durch den Filter 12 analysiert, um entweder positiv eine Sprachgruppe zu identifizieren oder um bestimmte Sprachgruppen von der weiteren Betrachtung auszuschließen.

- Der Filter 12 arbeitet, um Sprachgruppen für Eingangsnamen
25 auf der Basis eines vorgegebenen Regelsatzes auszufiltern. Die Regeln werden dem Filter 12 von einem Regel-Speicher, der später beschrieben wird, bereitgestellt.

- Jeder Eingangsname wird betrachtet, als aus einer Kette von
30 Graphemen zusammengesetzt zu sein. Einige Ketten innerhalb eines Eingangsnamens identifizieren (oder eliminieren) eindeutig eine Sprachgruppe für diesen Namen. Zum Beispiel identifiziert gemäß einer Regel die Kette BAUM positiv den Eingangsnamen als deutsch (z.B. TANNENBAUM). Gemäß einer
35 anderen Regel identifiziert die Kette MOTO am Ende eines Namens positiv die Sprachgruppe als japanisch (z.B. KAWAMOTO). Wenn es eine solche positive Identifikation gibt, werden der Eingangsname und die identifizierte Sprachgruppe (L TAG) direkt zum Buchstabe-zu-Klang-Abschnitt 20 gesendet,

der die geeigneten Phoneme der Sprachrealisierungseinheit 50 zur Verfügung stellt.

Der Filter 12 versucht ansonsten, so viele Sprachgruppen wie
 5 möglich von der weiteren Betrachtung zu eliminieren, wenn eine positive Identifikation nicht möglich ist. Dies erhöht die Wahrscheinlichkeitsgenauigkeit der verbleibenden Analyse des Eingangsnamens. Zum Beispiel liefert eine Filterregel, daß, wenn die Kette -B am Ende eines Namens ist, Namensgruppen wie
 10 zum Beispiel Japanisch, Slawisch, Französisch, Spanisch und Irisch von einer weiteren Betrachtung eliminiert werden können. Durch diese Eliminierung wird die nachfolgende Analyse zum Bestimmen der Ursprungs-Sprachgruppe für einen Eingangsnamen, der nicht positiv identifiziert wird, vereinfacht und verbessert.
 15

Unter der Annahme, daß keine Sprachgruppe positiv als die Ursprungs-Sprachgruppe durch den Filter 12 bestimmt werden kann, ist eine weitere Analyse notwendig. Diese wird durch
 20 den Trigramm-Analysierer 14 durchgeführt, der den Eingangsnamen und die Liste aller Sprachgruppen empfängt, die nicht durch den Filter 12 eliminiert wurden. Der Trigramm-Analysierer 14 analysiert die Kette aus Graphemen (den Eingangsnamen) syntaktisch in Trigrammen, die Graphem-Ketten
 25 sind, die drei Grapheme lang sind. Zum Beispiel wird die Graphem-Kette # SMITH # in den folgenden fünf Trigramme syntaktisch analysiert: # SM, SMI, MIT, ITH, TH # . Für die Trigramm-Analyse wird das Zwischenzeichen (Wortgrenze) als Graphem betrachtet. Deshalb ist die Anzahl der Trigramme
 30 immer die gleiche wie die Anzahl der Grapheme im Namen.

Die Wahrscheinlichkeit für jedes der Trigramme dafür, daß es von einer bestimmten Sprachgruppe ist, wird dem Trigramm-Analysierer 14 eingegeben. Diese Wahrscheinlichkeit, die
 35 durch eine Analyse einer Namendatenbank berechnet wird, wird als ein Eingang von einer Häufigkeitstabelle aus Trigrammen für jede Sprachgruppe empfangen, die nicht durch den Filter 12 eliminiert wurde. Das Gleiche wird auch für jedes der anderen Trigramme der Graphem-Kette ausgeführt.

Die nachfolgende (teilweise) Matrix zeigt Proben-Wahrscheinlichkeiten für den Nachnamen VITALE:

5		Li	Lj	Ln
	#VI	,0679	,4659	,2093
	VIT.	,0263	,4145	,0000
10	ITA	,0490	,7851	,0564
	TAL	,1013	,4422	,2384
	ALE	,0867	,2602	,2892
	LE#	,1884	,3181	,0688

15	Gesamtwahr- scheinlichkeit	,0866	,4477	,1437

In dem obenstehenden Feld ist L eine Sprachgruppe und n die Anzahl der Sprachgruppen, die durch den Filter 12 nicht eliminiert wurden. Das Trigramm #VI hat eine Wahrscheinlichkeit von 0,0679 dafür, daß es aus der Sprachgruppe Li ist, von 0,4659 dafür, daß es aus der Sprachgruppe Lj ist, und von 0,2093 dafür, daß es aus der Sprachgruppe Ln ist. Lj wird als die höchste Wahrscheinlichkeit herausgemittelt und somit ist die Sprachgruppe identifiziert.

Die Wahrscheinlichkeit für jedes der Trigramme der Graphem-Kette (Eingangsname) wird ähnlich dem Trigramm-Analysierer 14 eingegeben. Die Wahrscheinlichkeit für jedes Trigramm in einem Eingangsnamen wird für jede Sprachgruppe gemittelt. Dies gibt die Wahrscheinlichkeit des Eingangsnamens, der von einer bestimmten Sprachgruppe stammt, wieder. Die Wahrscheinlichkeit, daß die Graphem-Kette # VITALE # zu einer bestimmten Sprachgruppe gehört, wird als Vektor von Wahrscheinlichkeiten von der gesamten Wahrscheinlichkeitslinie erzeugt. Aus diesen Vektor von Wahrscheinlichkeiten können auch andere Größen, wie zum Beispiel die Standardabweichung und die Schwellenüberprüfung, berechnet werden. Dies stellt sicher,

daß ein einzelnes Trigramm nicht über der Gesamtwahrscheinlichkeit liegen kann, die Gesamtwahrscheinlichkeit bildet oder verzerrt.

- 5 Obwohl die illustrierte Ausführungsform Trigramme analysiert, kann der Analysierer 14 konfiguriert werden, um unterschiedlich lange Graphem-Ketten, wie zum Beispiel eine Kette mit zwei Graphemen oder vier Graphemen zu analysieren.
- 10 In dem obenstehenden Beispiel zeigt der Trigramm-Analysierer 14, daß die Sprachgruppe L_j die wahrscheinlichste Ursprungssprachgruppe für den gegebenen Eingangsnamen ist, da sie die höchste Wahrscheinlichkeit hat. Es ist diese wahrscheinlichste Sprachgruppe, die L TAG für den Eingangsnamen wird.
- 15 Das L TAG und der Eingangsname werden dann dem Buchstaben-zu-Klang-Abschnitt 20 zugesendet, um die Phoneme für den Eingang zu erzeugen. Die Filterregeln sind auf eine solche Art und Weise aufgebaut, daß eine Zweideutigkeit der Identifikation nicht möglich ist. Das heißt, daß eine Sprache nicht sowohl
20 eliminiert als auch positiv identifiziert werden kann, da ein Dominanz-Verhältnis derart angewendet wird, daß eine positive Identifikation über eine Eliminierungs-Regel bei dem unwahrscheinlichen Ereignis eines Konflikts dominiert.
- 25 Ähnlicherweise kann eine Sprachgruppe nicht positiv für mehr als eine Sprache identifiziert werden, da die Filterregeln einen geordneten Satz derart bilden, daß die erste positive Identifikation angewendet wird.
- 30 Das System kann eine bestimmte Sprachgruppe standardklassifizieren, wenn eines von zwei Schwellen-Kriterien eingehalten wird: (a) ein absolutes Schwellenüberprüfen tritt auf, wenn die höchste Wahrscheinlichkeit, die vom Trigramm-Analysierer 14 bestimmt
35 wird, unterhalb eines vorgegebenen Schwellenwerts T_i ist. Dies würde bedeuten, daß der Trigramm-Analysierer 14 aus den Sprachgruppen eine einzelne Sprachgruppe mit einem vernünftigen Vertrauens-Grad nicht bestimmen kann; (b) eine relative Schwellenwertüberprüfung tritt auf, wenn der

Unterschied in den Wahrscheinlichkeiten zwischen der Sprachgruppe, die mit der höchsten Wahrscheinlichkeit identifiziert wurde, und der Sprachgruppe, die mit der zweithöchsten Wahrscheinlichkeit identifiziert wurde, unterhalb eines Schwellenwertes T_j fällt, wie durch den Trigramm-Analysierer 14 bestimmt wird.

Die Standardklassifizierung in eine spezifizierte Sprachgruppe ist ein setzbarer Parameter. Zum Beispiel ist in einer Englisch sprechenden Umgebung eine Standardklassifizierung für eine englische Aussprache im allgemeinen der sicherste Weg, da ein Mensch, dem ein niedriges Vertrauensniveau gegeben ist, am wahrscheinlichsten auf die gattungsgemäße englische Aussprache des Eingangsnamens zurückgreifen wird. Der Wert der Standardklassifizierung als ein setzbarer Parameter besteht darin, daß die Standardklassifizierung in gewissen Situationen, zum Beispiel, wenn die Telefonnummer anzeigt, daß eine Telefonnummer in einer relativ homogenen ethnischen Nachbarschaft lokalisiert ist, geändert werden kann.

Wie zuvor erwähnt wurde, werden der Name und das Sprach-Zeichen (LTAG), das entweder durch den Filter 12 oder den Trigramm-Analysierer 14 zugesendet wird, von dem Buchstabe-zu-Klang-Regelabschnitt 20 empfangen. Der Buchstabe-zu-Klang-Regelabschnitt 20 ist konzeptionell in separate Blöcke für jede Sprachgruppe aufgeteilt. D.h., daß die Sprachgruppe (L_i) ihren eigenen Satz von Buchstabe-zu-Klang-Regeln hat, wie es auch die Sprachgruppe (L_j), die Sprachgruppe (L_k) usw. bis zur Sprachgruppe (L_n) haben.

Unter der Annahme, daß der Eingangsname ausreichend identifiziert wurde, so daß eine Standardklassifizierungsausprache nicht erzeugt wird, wird der Eingangsname zu dem geeigneten Sprachgruppe-Buchstabe-zu-Klang-Block 22_{i-n} gemäß dem Sprache-Kennzeichen gesendet, das mit dem Eingangsnamen verbunden ist.

In dem Buchstabe-zu-Klang-Regelabschnitt 20 sind die Regeln für die einzelnen Sprachgruppen-Blöcke 22 Untersätze eines größeren und komplexeren Satzes aus Buchstabe-zu-Klang-Regeln für andere Sprachgruppen, einschließlich Englisch. Ein

5 Buchstabe-zu-Klang-Block 22_i für eine spezifische Sprachgruppe L_i, die als Ursprungs-Sprachgruppe identifiziert wurde, versucht, mit der größten Graphem-Sequenz für eine Regel übereinzustimmen. Dies ist unterschiedlich zu dem Filter 12, der von oben nach unten und in dieser Ausführungsform von

10 rechts nach links nach der Kette aus Graphemen in einem Eingangsnamen sucht, der eine Filter-Regel erfüllt. Der Buchstabe-zu-Klang-Block 22_{i-n} für eine spezifische Sprache tastet die Graphem-Kette von links nach rechts oder von rechts nach links ab, wobei die illustrierte Ausführungsform

15 eine Rechts-nach-Links-Abtastung verwendet.

Ein Beispiel für die Buchstabe-zu-Klang-Regeln für einen spezifischen Block L_i kann für einen Namen wie zum Beispiel MAN-KIEWICZ betrachtet werden. Dieser Eingangsname würde als von

20 einer slawischen Sprachgruppe stammend identifiziert werden, die die höchste Wahrscheinlichkeit hat, und würde deshalb zu den slawischen Buchstabe-zu-Klang-Regelblock 22_i gesendet werden. In diesem Block 22_i hat die Graphem-Kette -WICZ eine Ausspracheregeln, um die korrekten segmentartigen Phoneme der

25 Kette bereitzustellen. Die Graphem-Kette -KIEWICZ hat jedoch auch eine Regel in dem slawischen Regelsatz. Da dies eine längere Graphem-Kette ist, würde diese Regel zuerst angewendet werden. Die segmentartigen Phoneme für alle übrigen Grapheme, die nicht einer sprachspezifischen Ausspracheregeln entsprechen, werden dann von dem allgemeinen Ausspracheblock

30 bestimmt. In diesem Beispiel würden die segmentartigen Phoneme für Grapheme M, A und N gemäß den allgemeinen Ausspracheregeln (getrennt) bestimmt werden. Der Buchstabe-zu-Klang-Block 22_i sendet die verketteten Phoneme sowohl für die

35 sprachsensitiven Graphem-Ketten als auch für die nicht-sprachsensitiven Graphem-Ketten zusammen der Sprachrealisierungseinheit 50 für die Aussprache zu.

Der Filter 12 enthält nicht alle größeren Ketten, die sprachspezifisch sind und die in den Buchstabe-zu-Klang-Regeln 20 enthalten sind. Die größeren Ketten werden nicht alle benötigt, da zum Beispiel die Kette -WICZ positiv einen Eingangs-

5 namen als vom Ursprung her slawisch identifizieren würde. Es besteht dann für die Filterregel für die Kette -KIEWICZ keine Notwendigkeit, da -WICZ ein Untersatz von -KIEWICZ ist und somit den Eingangsnamen identifizieren würde.

- 10 Das Buchstabe-zu-Klang-Modul gibt die Phoneme für Namen hauptsächlich in der Form von segmentartiger Phonem-Information aus. Der Ausgang der Buchstabe-zu-Klang-Regel-Blöcke 22_{i-n} dient als Eingang für die Betonungs-Abschnitte 24_{i-n} . Diese Betonungs-Abschnitte 24_{i-n} nehmen das LTAG zusammen mit
- 15 den Phonemen, die durch einzelne Buchstabe-zu-Klang-Regel-Blöcke 22_{i-n} erzeugt werden, und gibt eine komplette Phonem-Kette aus, die sowohl segmentartige Phoneme (von Buchstabe-zu-Klang-Regel-Blöcken 22_{i-n}) als auch die korrekten Betonungsmuster für diese Sprache enthalten. Wenn zum Bei-
- 20 spiel die Sprache, die für den Namen VITALE identifiziert wurde, Italienisch war und der Buchstabe-zu-Klang-Regel-Block 22 die Phonem-Kette [vitali] liefert, dann würde der Betonungsabschnitt 24_i die Betonung auf die vorletzte Silbe legen, so daß die endgültige phonetische Kette [vitáli] sein
- 25 würde.

- Es sollte beachtet werden, daß tatsächliche Regeln, die in dem Filter 12, in dem Buchstabe-zu-Klang-Abschnitt 20 und in dem Betonungsabschnitten 24_{i-n} verwendet werden, Regeln sind,
- 30 die entweder bekannt sind oder einfach von einem Fachmann in Linguistik erhalten werden können.

- Das oben beschriebene System kann als Frontend-Prozessor für eine Sprachrealisierungseinheit 50 betrachtet werden. Die
- 35 Sprachrealisierungseinheit 50 kann eine kommerziell erhältliche Einheit zum Erzeugen von menschlicher Sprache aus graphen-

mischer und phonemischer Eingabe sein. Der Synthesizer kann auf Phonemen basieren oder auf einer anderen Klangeinheit, zum Beispiel Diphon oder Halbsilbe. Der Synthesizer kann auch eine andere Sprache als Englisch synthetisieren.

5

Figur 2 zeigt einen Block 60 für Sprachgruppen-Identifikation und phonetische Realisierung als Teil eines Systems. Der Block 60 für Sprachgruppen-Identifikation und phonetische Realisierung besteht aus den Funktionsblöcken, die in der Figur 1 gezeigt werden. Wie gezeigt wird, ist der Eingang für den Block 60 für Sprachidentifikation und phonetische Realisierung der Name, die Filterregeln und die Trigramm-Wahrscheinlichkeiten. Der Ausgang ist der Name, das Sprachkennzeichen und die Phoneme, die zu der Sprachrealisierungseinheit 50 gesendet werden. Es sollte beachtet werden, daß Phoneme in diesen Zusammenhang irgend ein Alphabet aus Klang-Symbolen, einschließlich Diphonen und Halbsilben, bedeutet.

Das System gemäß der Figur 2 markiert Graphem-Ketten als zu einer bestimmten Sprachgruppe zugehörig. Der Sprach-Identifizierer wird zum Vorfiltern einer neuen Datenbank verwendet, um die Wahrscheinlichkeits-Tabelle für eine bestimmte Datenbank zu verfeinern. Der Analyseblock 62 empfängt als Eingänge den Namen und das Sprachkennzeichen und Statistiken von dem Block 60 für Sprachidentifikation und phonetische Realisierung. Der Analyseblock übernimmt diese Informationen und gibt den Namen und das Sprachkennzeichen an eine Master-Sprachdatei 64 aus und erzeugt Regeln für einen Filterregel-Speicher 68. Auf diese Art und Weise wird die Datenbank des Systems erweitert, wenn neue Eingangsnamen verarbeitet werden, so daß künftige Eingangsnamen einfacher verarbeitet werden können. Der Filterregel-Speicher 68 stellt die Filterregeln für den Filter 12 und für den Block 60 für Sprachidentifikation und phonetische Realisierung bereit.

Die Master-Datei enthält alle Graphem-Ketten und ihre Sprachgruppen-Kennzeichen. Dieser Block 64 wird durch den Analyseblock 62 erzeugt. Die Trigramm-Wahrscheinlichkeiten werden in einer Datenstruktur 66 angeordnet, die für die

5 Einfachheit des Suchens für ein gegebenes Eingangs-Trigramm ausgelegt ist. Die illustrierte Ausführungsform verwendet zum Beispiel eine N-tiefe dreidimensionale Matrix, wobei n die Anzahl der Sprachgruppen ist.

10 Trigramm-Wahrscheinlichkeits-Tabellen werden aus der Master-Datei berechnet, indem der nachfolgende Algorithmus verwendet wird:

15 Berechne die Gesamtzahl von Auftritten jedes Trigramms für alle Sprachgruppen L (1-N);

für alle Graphem-Ketten S in L

für alle Trigramme T in S

wenn (count [T][L] = 0)

uniq [L] + = 1

20 count [T][L] + = 1

Für alle möglichen Trigramme T im Master

sum = 0

für alle Sprachgruppen L

25 sum + = count [T][L]/uniq [L]

für alle Sprachgruppen L

wenn sum >0, prob[T][L]=count [T][L]/uniq [L]/sum

ansonsten prob [T][L] = 0,0;

Die Trigramm-Häufigkeitstabelle, die zuvor erwähnt wurde, kann als dreidimensionales Feld von Trigrammen, Sprachgruppen und Häufigkeiten betrachtet werden. Häufigkeiten bedeuten den Prozentsatz an Auftreten jener Trigramm-Sequenzen für die jeweiligen Sprachgruppen auf der Basis einer großen Probe von Namen. Die Wahrscheinlichkeit eines Trigramms dafür, daß es ein Teilnehmer einer bestimmten Sprachgruppe ist, kann auf einer Anzahl von Arten abgelitten werden. In dieser Ausführungsform wird die Wahrscheinlichkeit eines Trigramms dafür, daß es ein Teilnehmer einer bestimmten Sprachgruppe ist, von dem gut bekannten Bayes-Theorem aus gemäß der Formel abgelitten, die untenstehend angegeben ist:

Die Bayes-Regel sagt, daß die Wahrscheinlichkeit, daß Bj bei gegebenem A auftritt, $P(Bj|A)$, gleich

$$P(Bj|A) = \frac{P(A|Bj)P(Bj)}{\sum_i P(A|Bi)P(Bi)} \quad \text{ist.}$$

Genauer für das Problem beträgt die Wahrscheinlichkeit für eine Sprachgruppe bei gegebenem Trigramm T gleich $P(Li|T)$, wobei

$$P(Li|T) = \frac{P(T|Li)P(Li)}{\sum_k P(T|L_k)P(L_k)}$$

weiterhin analysierend

$$P(T|Li) = \frac{X}{Y}$$

wobei X = Anzahl, wie oft das Zeichen T in der Sprachgruppe Li auftritt;

Y = Anzahl der einmalig auftretenden Zeichen
in der Sprachgruppe L_i

$$P(L_i) = \frac{1}{N} \text{ immer}$$

5

wobei N = die Anzahl der Sprachgruppen (nicht überlappend)

$$10 \quad P(L_i | T) = \frac{\frac{P(T|L_i)}{N}}{\sum_{k=1}^N \frac{P(T|L_k)}{N}} = \frac{P(T|L_i)}{\sum_{k=1}^N P(T|L_k)}$$

15 Die endgültige Tabelle hat dann vier Dimensionen: eine für jedes Graphem des Trigramms und eine für die Sprachgruppe.

Die Trigramm-Wahrscheinlichkeiten werden, wie von dem Block 66 berechnet, zu dem Block 60 für Sprachidentifikation und phonetische Realisierung und insbesondere zu dem Trigramm-Analysierer 14 gesendet, der den Wahrscheinlichkeitsvektor dafür erzeugt, daß die Graphem-Kette zu einer bestimmten Sprachgruppe gehört.

25 Unter Verwendung des oben beschriebenen Systems können Namen genauer ausgesprochen werden. Weitere Entwicklungen, wie zum Beispiel das Verwenden des Vornamens in Verbindung mit dem Nachnamen, um den Nachnamen genauer auszusprechen, sind mit zu betrachten. Dies würde eine Ausdehnung der existierenden Wissens-Bank und der Regelsätze umfassen.

30

Ansprüche

- 5 1. Verfahren zum positiven Identifizieren oder Eliminieren einer Sprachgruppe ($L_1 \dots L_n$) als eine Ursprungs-Sprachgruppe für ein gegebenes Wort, das aufweist:
Vergleichen von Unterketten aus Graphemen eines Eingangswortes mit einem gespeicherten Satz von Filterregeln, bis entweder
10 eine Übereinstimmung einer der Unterketten mit einer der Filterregeln positiv eine Sprachgruppe identifiziert oder bis irgendeine Sprachgruppe eliminiert wird, wenn eine Übereinstimmung einer der Unterketten mit einer der Filterregeln anzeigt, daß eine Sprachgruppe von der Betrachtung
15 als eine Ursprungs-Sprachgruppe für das Eingangswort eliminiert werden kann; und
Erzeugen einer Liste möglicher, nicht eliminierter Ursprungs-Sprachgruppen, wenn keine Sprachgruppe positiv als die Ursprungs-Sprachgruppe identifiziert wird, oder
20 Anzeigen der Ursprungs-Sprachgruppe, wenn die Ursprungs-Sprachgruppe positiv identifiziert wird.
2. Verfahren, wie in Anspruch 1 beansprucht, wobei der Vergleichsschritt den Schritt des Durchsuchens der Filterregeln
25 von oben nach unten und von rechts nach links enthält.
3. Verfahren, wie in Anspruch 1 beansprucht, wobei der Vergleichsschritt den Schritt des Durchsuchens der Filterregeln nach Sprachgruppe und nach Graphem innerhalb jeder Sprachgruppe
30 enthält.
4. Verfahren zum Erzeugen korrekter Phoneme für ein gegebenes Eingangswort gemäß einer Sprachgruppe von Ursprüngen des Eingangswortes, wobei das Verfahren aufweist:
35 Filtern des Eingangswortes in einem Filter (12), um eine Ursprungs-Sprachgruppe für das Eingangswort zu identifizieren

oder um zumindest eine Ursprungs-Sprachgruppe für das Eingangswort zu eliminieren;

Senden des Eingangswortes und einer Sprach-Kennzeichnung, die eine Ursprungs-Sprachgruppe für das Eingangswort anzeigt, von dem Filter zu einem Buchstabe-zu-Klang-Modul (22), der Buchstabe-zu-Klang-Regeln enthält, wenn der Filter positiv eine Ursprungs-Sprachgruppe für das Eingangswort identifiziert;

Senden des Eingangswortes und jeder nicht eliminierten Sprachgruppe von dem Filter aus zu einem Graphem-Analysierer (14), wenn eine Ursprungs-Sprachgruppe für das Eingangswort nicht positiv durch den Filter identifiziert wird;

Erzeugen einer wahrscheinlichsten Ursprungs-Sprachgruppe für das Eingangswort, indem Grapheme in dem Eingangswort analysiert werden;

Senden des Eingangswortes und der wahrscheinlichsten Ursprungs-Sprachgruppe zu einem Untersatz des Buchstabe-zu-Klang-Moduls entsprechend der wahrscheinlichsten Sprachgruppe;

Erzeugen von segmentartigen Phonemen für das Eingangswort in dem Untersatz des Buchstabe-zu-Klang-Moduls;

Senden der segmentartigen Phoneme und der Sprach-Kennzeichnung von dem Buchstabe-zu-Klang-Modul zu einem Betonungs-Zuordnungsabschnitt (24);

Erzeugen einer Betonungs-Zuordnungs-Information für das Eingangswort in dem Betonungs-Zuordnungsabschnitt; und

Senden der segmentartigen Phoneme und der Betonungs-Zuordnungs-Information zu einer Sprachrealisierungseinheit (50).

5. Verfahren, wie in Anspruch 4 beansprucht, wobei die Grapheme Trigramme sind.

6. Verfahren, wie in Anspruch 4 oder 5 beansprucht, wobei der Schritt des Erzeugens einer wahrscheinlichsten Ursprungs-Sprachgruppe den Schritt des Ermitteln von Wahrscheinlichkeiten der Grapheme für ein Eingangswort, das aus einer bestimmten Sprachgruppe ist, unter Verwendung der Bayes-Regel enthält.

7. Verfahren, wie in Anspruch 4, 5 oder 6 beansprucht, das weiterhin den Schritt des Einstellens auf eine allgemeine Aussprache aufweist, wenn der Schritt des Erzeugens einer
 5 wahrscheinlichsten Ursprungs-Sprachgruppe eine wahrscheinlichste Ursprungs-Sprachgruppe mit einer Wahrscheinlichkeit unterhalb eines vorgegebenen Schwellenniveaus erzeugt.
- 10 8. Verfahren, wie in Anspruch 4, 5, 6 oder 7 beansprucht, das weiterhin den Schritt des Einstellens auf eine allgemeine Aussprache aufweist, wenn der Schritt des Erzeugens einer wahrscheinlichsten Ursprungs-Sprachgruppe eine
 15 wahrscheinlichste Ursprungs-Sprachgruppe mit einer Wahrscheinlichkeit erzeugt, die nicht größer um einen bestimmten Wert als eine Wahrscheinlichkeit der nächsten wahrscheinlichsten Ursprungs-Sprachgruppe ist.
- 20 9. Verfahren, wie in irgendeinem der Ansprüche 4 bis 8 beansprucht, das ein erstes Durchsuchen eines Verzeichnisses (10) nach einem Eintrag entsprechend dem Eingangswort enthält, wobei jeder Eintrag ein Wort und Phoneme für dieses Wort enthält; und
 25 und ein Senden eines Eintrags zu der Sprachrealisierungseinheit für die Aussprache enthält, wenn die Verzeichnis-Durchsuchung diesen Eintrag entsprechend den Eingangswörtern ergibt.
- 30 10. Vorrichtung zum positiven Identifizieren oder Eliminieren einer Sprachgruppe ($L_1 \dots L_n$) als eine Ursprungs-Sprachgruppe für ein gegebenes Wort, die aufweist:
 35 einen Filterregel-Speicher (68), der einen Satz mit Filterregeln speichert, wobei ein erster Untersatz der Filterregeln positiv eine Sprachgruppe identifiziert und wobei ein zweiter Satz der Filterregeln eine Sprachgruppe eliminiert;
 einen Vergleicher (12), der Unterketten der Grapheme eines Eingangswortes mit dem ersten Untersatz und dem zweiten

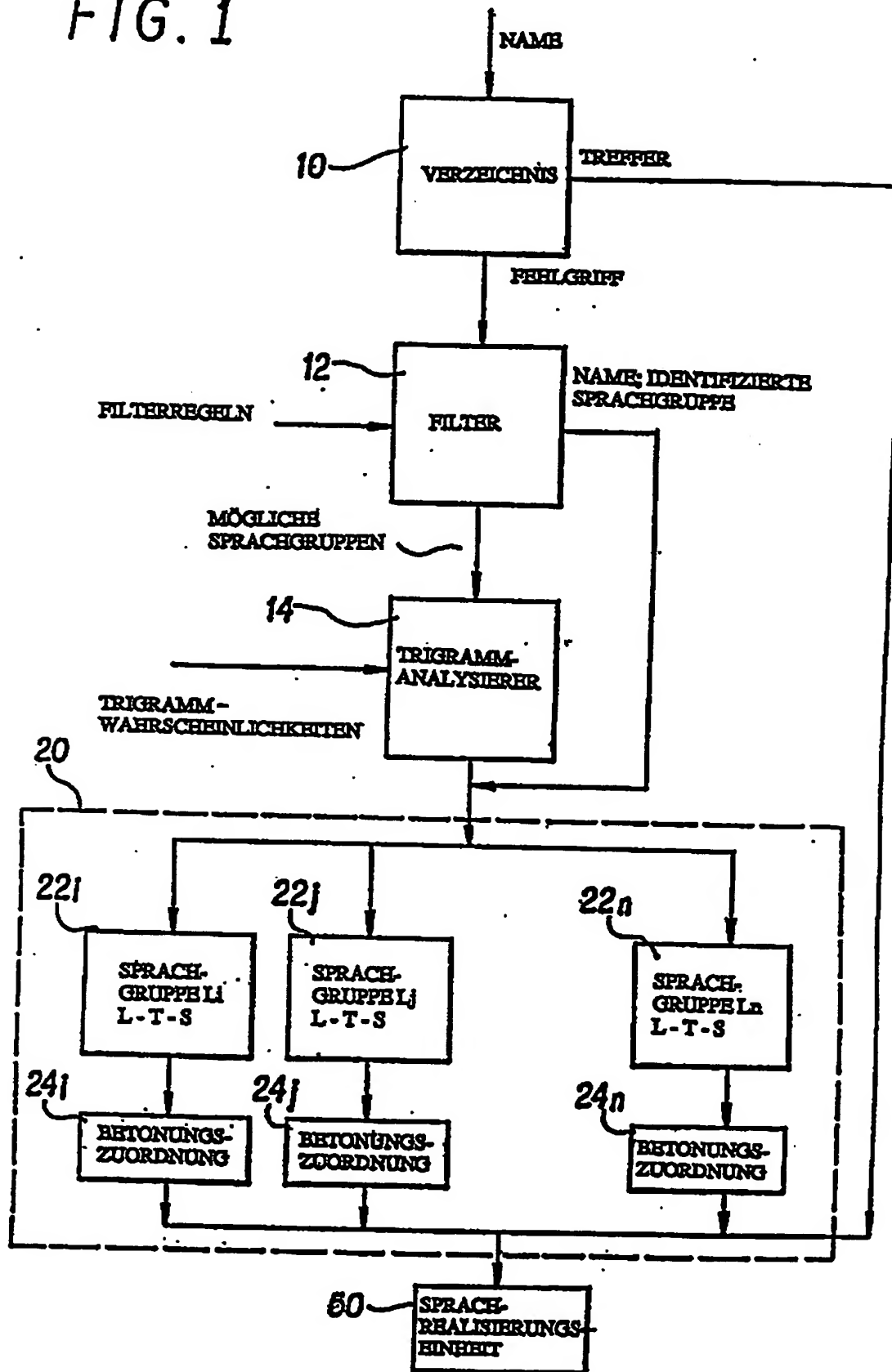
Untersatz der Filterregeln vergleicht, bis eine Übereinstimmung einer der Unterketten mit einer des ersten Untersatzes von Filterregeln positiv eine Sprachgruppe identifiziert oder eine Sprachgruppe eliminiert, wenn eine Übereinstimmung einer der
5 Unterketten mit einer des zweiten Untersatzes der Filterregeln anzeigt, daß eine Sprachgruppe von der Betrachtung als eine Ursprungs-Sprachgruppe für das Eingangswort eliminiert ist; und einen Ausgang, der eine Liste möglicher Ursprungs-Sprachgruppen erzeugt, wenn keine Sprachgruppe als Ursprungs-Sprachgruppe
10 positiv identifiziert wird, und der eine Anzeige der Ursprungs-Sprachgruppe erzeugt, wenn die Ursprungs-Sprachgruppe positiv identifiziert wird.

11. Vorrichtung, wie in Anspruch 10 beansprucht, die einen
15 Analysierer (14) zum Berechnen der wahrscheinlichsten Ursprungs-Sprachgruppe für die Grapheme, die von dem Ausgang aus empfangen werden, in dem gegebenen Wort für jede Sprache enthält, die nicht durch den zweiten Untersatz der Filterregeln, eliminiert wird.

20 12. Vorrichtung, wie in Anspruch 11 beansprucht, in der der Analysierer Grapheme in dem gegebenen Wort analysiert, die in Trigrammen angeordnet sind.

1/2

FIG. 1



2/2

FIG. 2

